

И. Б. САЕНКО, И. Б. ПАРАЩУК, И. В. КОТЕНКО
Санкт-Петербургский институт информатики
и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург

ТЕОРЕТИЧЕСКИЙ И ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ КОМПОНЕНТА УСТРАНЕНИЯ НЕОПРЕДЕЛЕННОСТИ ОЦЕНКИ И КАТЕГОРИЗАЦИИ СМЫСЛОВОГО НАПОЛНЕНИЯ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ

Проведен анализ компонента устранения неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов для системы интеллектуальной аналитической обработки цифрового сетевого контента в целях защиты от нежелательной, сомнительной и вредоносной информации. Этот анализ проведен с точки зрения экспертной оценки качества программных средств, опираясь на сформулированные и обоснованные показатели качества, с учетом функциональных и иных требований к данному компоненту. Определены обобщенный интегральный показатель качества и общий класс качества компонента.

Введение. Важной задачей в рамках создания системы интеллектуальной аналитической обработки цифрового сетевого контента (ЦСК) в целях защиты пользователей инфотелекоммуникационных сетей от нежелательной, сомнительной и вредоносной информации, является теоретическая и экспериментальная оценка методов, моделей, методик, алгоритмов и программных прототипов компонентов этой системы.

Одно из ключевых мест в структуре современной системы интеллектуальной аналитической обработки ЦСК, наряду с компонентами сбора и предварительной обработки сетевых информационных объектов, многоаспектной оценки и категоризации смыслового наполнения (СН) информационных объектов, обеспечения своевременности многоуровневого и многомодульного анализа информационных объектов, адаптации и переобучения, а также компонентами выработки (выбора) мер противодействия и реализации визуальных интерфейсов, занимает компонент устранения неполноты и противоречивости оценки и категоризации СН информационных объектов, построенный на основе использования методов, алгоритмов и программных процедур обработки неполных, противоречивых и нечетких знаний.

Сущность проблемы заключается в том, что никакую научную, исследовательскую и практическую работу по созданию и подготовке к внедрению в систему интеллектуальной аналитической обработки ЦСК компонента устранения неполноты и противоречивости оценки и категоризации СН информационных объектов, нельзя считать завершенной, пока не реализован важнейший комплекс задач – проверка качества данного компонента. Тем более, что речь идет об оригинальном компоненте, способном работать в условиях различного вида неопределенности исходных данных, способном устранить как нечеткость, так и неполноту и противоречивость оценки и категоризации СН информационных объектов. Существующим традиционным подходом к решению подобной проблемы является теоретический и экспериментальный анализ данного компонента в рамках интеллектуальной аналитической обработки ЦСК в целях защиты от нежелательной, сомнительной и вредоносной информации.

Предлагаемый доклад посвящен описанию особенностей и результатов реализации этапов методологии теоретического и экспериментального анализа для конкретного элемента системы интеллектуальной аналитической обработки ЦСК – компонента устранения неполноты и противоречивости оценки и категоризации СН информационных объектов.

Особенности этапов, показатели и результаты теоретического и экспериментального анализа компонента устранения неполноты и противоречивости. Задачу, стоящую перед нами можно сформулировать следующим образом. Сервисы интеллектуальной аналитической обработки ЦСК, с точки зрения оценки качества представляют собой программные средства. По этой причине методология теоретической и экспериментальной оценки методов, моделей, методик, алгоритмов и программных прототипов интеллектуальной аналитической обработки ЦСК должна основываться на имеющихся отечественных и международных стандартах в области оценки качества программных средств.

Основные понятия, показатели качества и общий подход к анализу. С учетом того факта, что анализировать принято конечный продукт разработки – программные средства (ПС), рассмотрим основные понятия [1]: «качество ПС» – совокупность свойств ПС, которые обуславливают его пригодность удовлетворять заданные или подразумеваемые потребности в соответствии с его назначением; «свойство ПС» – отличительная особенность ПС, которая может проявляться при его создании, использовании, анализе или изменении; «критерий оценки (качества ПС)» – совокупность принятых в установленном порядке правил и условий, с помощью которых устанавливается приемлемость в целом качества программного средства.

С учетом этого обычно определяют понятие «уровень пригодности (программного средства)» – степень удовлетворения потребностей, представленная посредством конкретного набора значений характеристик качества ПС. В соответствии со стандартом [2] понятие «оценка качества программного средства» раскрывается как совокупность следующих операций: выбор номенклатуры показателей качества оцениваемого программного средства; определение значений этих показателей; сравнение их с базовыми значениями. Исходя из этого определения, можно сделать вывод, что основными задачами, которые подлежат решению при оценке методов, моделей, методик, алгоритмов и программного прототипа компонента устранения неполноты и противоречивости оценки и категоризации СН информационных объектов в рамках системы интеллектуальной аналитической обработки ЦСК, являются: определение номенклатуры показателей качества и определение методики расчета значений этих показателей.

Принято определять шесть ключевых характеристик [3], которые с минимальным дублированием, описывают качество ПС и образуют основу для дальнейшего уточнения и описания качества программного обеспечения: функциональные возможности; надежность; практичность; эффективность; сопровождаемость и мобильность. Они, в свою очередь, являются комплексными показателями качества ПС и включают [3]: функциональные возможности – пригодность, правильность, способность к взаимодействию, согласованность, защищенность; надежность – стабильность, устойчивость к ошибке, восстанавливаемость; практичность – понятность, обучаемость и простота использования; эффективность – характер изменения во времени и характер изменения ресурсов; сопровождаемость – анализируемость, изменяемость, устойчивость и тестируемость; мобильность – адаптируемость и взаимозаменяемость.

Общая методика анализа качества ПС опирается на мнения экспертов и включает в себя следующие шаги. На первом шаге осуществляется определение множества экспертов и множества возможных оцениваемых показателей качества. На втором этапе каждый эксперт осуществляет присвоение каждому из оцениваемых показателей баллов от 0 до 5 в соответствии с заданными критериями оценивания факторов качества в баллах. Затем осуществляется вычисление интегрального показателя качества для каждого эксперта (с учетом заранее оговоренных весовых коэффициентов). Потом, в зависимости от значения интегрального показателя качества определяется обобщенный интегральный показатель качества и класс качества программного средства. Значение обобщенного интегрального показателя качества рассчитывается как средневзвешенное значение с учетом нормированных весовых коэффициентов, характеризующих компетентность экспертов. В итоге, определяется класс качества ПС в соответствии со значениями обобщенного интегрального показателя качества (отличный (0–0,25), хороший (0,25–0,5), удовлетворительный (0,5–0,75), неудовлетворительный (0,75–1,0)).

Содержание и результаты анализа компонента устранения неопределенности оценки и категоризации СН информационных объектов. Для решения задачи анализа составлен перечень функциональных требований к компоненту устранения неопределенности оценки и категоризации смыслового наполнения информационных объектов, он детализирован под эту конкретную задачу, обозначается буквой латинского алфавита, опирается на рассмотренные ранее комплексные показатели качества и включает: показатели функциональных возможностей (определяются на основе функционального тестирования) – поддержка взаимодействия с другими компонентами (*A*), возможность предварительной обработки (*B*), ввод и обработка экспертной информации (*C*), решение задачи устранения неопределенности на основе вычисления систем нечетких множеств и реализация алгоритмов нейросетевой идентификации (*D*), вывод данных (*E*); показатели эффективности (определяются на основе нагрузочного тестирования) – временные показатели устранения неоднозначности (нечеткости) оценки и категоризации (*F*) и временные показатели устранения недостаточности (неполноты, противоречивости)

атрибутов СН информационных объектов (G); показатели сопровождаемости (определяются на основе статического анализа) – гибкость списка (данных) результатов предварительного анализа (H), гибкость входного формата экспертной информации (I) и гибкость формы отчетов (J); показатель мобильности – кроссплатформенность (K).

Получена статистика по экспертным оценкам вышеперечисленных показателей, которые были сделаны 10 экспертами. Было выполнено распределение весовых коэффициентов для оцениваемых показателей ($A, B, C, D, E, F, G, H, I, J$ и K). Для этой цели каждый эксперт упорядочивал показатели по убыванию их приоритета. Произведены вычисления интегральных показателей качества для всех экспертов. Затем эти коэффициенты были нормализованы с учетом компетентности экспертов и, в зависимости от значения интегрального показателя качества, был определен обобщенный интегральный показатель качества и класс качества программного средства. Расчетное значение обобщенного интегрального показателя качества составило 0,254.

Таким образом, согласно ранее введенным критериям, класс качества компонента устранения неопределенности оценки и категоризации смыслового наполнения информационных объектов на основе использования методов обработки неполных, противоречивых и нечетких знаний равен 1, что соответствует по уровню ранжирования качества, «хорошему» уровню.

Заключение. Таким образом, в соответствии с принятой методологией, проведен теоретический и экспериментальный анализ компонента устранения неполноты и противоречивости оценки и категоризации СН информационных объектов в рамках интеллектуальной аналитической обработки ЦСК в целях защиты от нежелательной, сомнительной и вредоносной информации. Этот анализ проведен с точки зрения экспертной оценки качества программных средств, опираясь на сформулированные и обоснованные показатели качества, с учетом функциональных и иных требований к данному компоненту. Определены обобщенный интегральный показатель качества и класс качества компонента устранения неопределенности оценки и категоризации смыслового наполнения информационных объектов. Возможным направлением дальнейших исследований может быть теоретический и экспериментальный анализ методов, моделей, методик, алгоритмов и программного прототипа компонента выработки и выбора мер противодействия нежелательной, сомнительной и вредоносной информации.

Работа проводилась при поддержке гранта РФФИ № 18-11-00302

ЛИТЕРАТУРА

1. International Standard ISO/IEC TR 9126-4:2004. Software engineering. Product quality. Part 4: Quality in use metrics. JTC 1/SC 7. 2004. 66 p.
2. International Standard ISO/IEC 25010:2011(E). Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE). System and software quality models. JTC 1/SC 7. 2011. 44 p.
3. Marinescu C. Quality Assessment of Enterprise Software Systems. *Published in: 2006 13th Working Conference on Reverse Engineering*. 2006. IEEE Xplore. P. 28–36.

I.B. Saenko, I.B. Parashchuk, I.V. Kotenko, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg

Theoretical and experimental analysis of the component of removing uncertainty evaluating and categorizing the semantic content of information objects

The analysis of the component to eliminate the incompleteness and inconsistency of the assessment and categorization of the semantic content of information objects for the system of intelligent analytical processing of digital network content in order to protect against unwanted, doubtful and malicious information is carried out. This analysis was carried out from the point of view of expert evaluation of the quality of software tools, relying on formulated and reasonable quality indicators, taking into account the functional and other requirements for this component. The generalized integral indicator of quality and the general class of component quality are determined.